

Predictive models of electric vehicle adoption in the United States: Charging ahead with renewable energy

Arnold Kamis^{a,*}, Preethi Susan Abraham^b

^a Associate Professor of Data Analytics, International Business School, Brandeis University, Sachar International Center, Office 107-004, Mailstop 032, 415 South St., Waltham, MA 02453, United States

^b Brandeis University, United States

ARTICLE INFO

Keywords:

Electric vehicles
Education
Solar power
Air quality
Gradient boosted trees
Regression

ABSTRACT

In this paper, we develop several predictive models pertaining to Electric Vehicles in the United States. We set out to understand three phenomena: Public Charging Units, Electric Vehicle Jobs, and Electric Vehicle Registrations. To model them, we include variables from various categories – demographics, economics, education, environment, finance, geographics, public health, and technographics – in the timespan of 2010–2020. We integrate data from multiple data sources and use them to understand, explain, and predict the Electric Vehicle phenomena combining state and county level data. We obtain three random effects linear regression models having variance explained of 51.5–62.4%. We also fit several machine learning models to improve the accuracy of the models, highlighting nonlinearities, to provide additional insights. The overall predictive accuracy of each Gradient Boosted Tree model is far superior to that of each linear regression model and significantly better than the other machine learning models. Our core contribution is that, spanning all three phenomena, the three most important predictor variables are solar generation of electricity, high school graduation rates, and air quality. We interpret the models and discuss their implications for research and practice.

Introduction

The climate crisis has implicated several economic sectors. They are 1) electricity, heat production, and other energy, 2) agriculture, forestry, other land use, 3) industry, 4) cities and buildings, and 5) transportation. To mitigate and reverse carbon emissions in all these sectors is the existential challenge of our day. Fortunately, our efforts at all scales – from national to local – are having a noticeable impact. Government policies, laws, and investments are slowing the increase in greenhouse gas emissions. Individual choices, social norms, business practices, technological advancements, and private investments are making progress as well. The focus of this paper is in sector 5) transportation, specifically electric vehicles, which have evolved rapidly in recent years (see Fig. 1). This paper has two research aims: 1) develop predictive models of Electric Vehicle adoption to identify the factors that increase or decrease adoption, and 2) compare and contrast the best linear statistical models against machine learning models to determine whether there are significant non-linearities at hand. Our core contribution is the identification of the most important predictor variables from our best, most predictive models.

In 1832, Robert Anderson developed the first Electric Vehicle (EV) in Scotland, and in 1896, William Morrison created the first successful Electric Vehicle in the United States. In 1997, Toyota introduced the first mass-produced hybrid, the Prius, and released it worldwide in 2000, ushering in the beginning of the modern era for Electric Vehicles. In 2006, Tesla Motors raised the bar further by promising a luxury electric sports car with 200 + miles of range.

In 2010, GM introduced the Chevy Volt, the first commercially available plug-in hybrid. Also in 2010, Nissan introduced the all-electric LEAF. The Electric Vehicle revolution was underway. The current era comprises three types of electric vehicles: the Hybrid Electric Vehicle, the Plug-In Hybrid Electric Vehicle, and the 100 % Electric Vehicle (also known as All-Electric Vehicle or Battery Electric Vehicle), as illustrated in Fig. 2. The traditional Internal Combustion Engine (ICE) is being augmented and replaced.

The public's interest in Electric Vehicles, which we define as encompassing all three types (hybrid, plug-in hybrid, and 100 % electric), has increased as Electric Vehicle batteries have increased in capacity and public charging units have become more widely available. The overall argument for Electric Vehicles has improved. The rising

* Corresponding author.

E-mail address: akamis@brandeis.edu (A. Kamis).

price of gasoline in 2022, caused in part by the Russian invasion of Ukraine, was exacerbated by supply chain problems, in general, and semiconductor chip shortages in the automotive industry, in particular. The public's interest in Electric Vehicles, which have attractive economies of fueling/charging, has been intensifying. In 2022, nearly all automakers added electrification to their offerings if they had not done so already.

Electric vehicles are increasingly attractive in many respects. They are new, high performance (strong torque), powered by electricity and/or gasoline, and cheaper to maintain than ICE cars. Nevertheless, there are obstacles that prevent people from buying an Electric Vehicle or trading in their ICE for an Electric Vehicle. The most common obstacle could be Range Anxiety: the fear or concern that an Electric Vehicle driver has about running out of battery power before reaching their destination or a charging station. This would be analogous to running out of gasoline with an ICE, but worse, since portable charging units are currently impractical (Afshar et al., 2021). Range Anxiety is particularly acute for a ride-sharing or taxicab driver (Hao et al., 2020). Could the main obstacle be the high initial purchase price of an Electric Vehicle (Knez et al., 2014), or are they available at all price points with most people eligible for financial incentives? There must be other factors apart from range anxiety and purchase price influencing the consumer's purchase/no purchase decision, but what are they and how can we model the totality of factors in a predictive model?

The rest of the paper is organized as follows. We first review literature in factors influencing Electric Vehicle adoption. We then describe our methods, including data sources for the variables of interest and descriptive statistics. We describe the results of our linear regression models and machine learning models, and then discuss our interpretation of the best models. We conclude by describing the overall research contributions, as well as limitations and future directions.

Literature summary

Fig. 3 summarizes the variables found commonly in the literature of Electric Vehicle adoption in the US and around the world, centered on public charging units, jobs, and registrations. The other twelve variables are either aspects of the potential Electric Vehicle buyer (e.g., demographics, education) or potential impacts of Electric Vehicles (e.g., public health, environment).

We group key findings from papers across the influential factors into five factor clusters.

Incentives: Financial incentives, e.g., tax credits, vary by geography and by perceived importance relative to EV range, depending on whether it is for personal vs. business use (Mersky et al., 2016) and also by specific make/model of the Electric Vehicle (Jenn et al., 2018). Availability of private parking (Patt et al., 2019), cost of electricity (Cunha et al., 2016), and access to the High-Occupancy Vehicle lane (Javid & Nejat, 2017) are also factors.

Environmental: Increasingly, people are putting solar panels on their roofs. Not only is this done to lower the resident's utility bill by purchasing less electricity from utility companies (Kaschub et al., 2016), but also to protect the environment (Noori et al., 2015). The propensity to

do so depends on the marginal cost of electricity (Onat et al., 2015), incentives to defray to cost of installations, and the resident's attitude toward the environment (Xu et al., 2021).

Public Health: Fewer tailpipe emissions can improve air quality, which results in better respiratory health. In Shanghai, China as well as in France and the US, hundreds of lives have been saved, with significant reductions in asthma and lost school days (Andre et al., 2020; Hu et al., 2021; Pan et al., 2021).

Geographics: In different regions, states, and cities, there are different factors influencing EV adoption, e.g., population density, GDP, and attitude towards automobile technology (Pan et al., 2021; Pan et al., 2019). Furthermore, different geographies have different mixes of vehicles (Lévy et al., 2017) and different summer/winter seasons. Charging time increases and Electric Vehicle efficiency decreases under extreme temperatures, particularly when cooling/heating the car cabin temperature (Jung, 2020).

Other Categories: Gender, age, and occupation are predictive factors, and people buy (or do not buy) Electric Vehicles for different reasons in the US and China (Zhang et al., 2017). The general attitude toward technological innovation varies along different social and socio-economic strata in the European Union (Christidis & Focas, 2019; Omahne et al., 2021). Different governments looking to phase out ICE vehicles can guide their citizens and their automakers, if any, with different climate, tax, energy, technology, and health policies, depending on their political and economic policies (Meckling & Nahm, 2019; Stokes & Breetz, 2018). At higher socio-economic levels, education levels and income are significant factors (Holland et al., 2018; Sovacool et al., 2018).

Ultimately, the proof of Electric Vehicle adoption is indicated by Electric Vehicle registrations. Our goal is to understand, explain, and predict Electric Vehicle registrations overall, in addition to two narrower phenomena: Public Charging Units and Electric Vehicle jobs. We define Electric Vehicle job as any job, occupation, or career related to electric vehicles. These jobs include manufacturing and assembly, research and development, charging infrastructure, sales and marketing, maintenance and repair, storage (e.g., battery) manufacturing and technology, software and controls, fleet management, policy and regulation, and environmental and sustainability analysis. The three phenomena of interest are our response variables:

- Public charging units. Range Anxiety is a significant factor in the minds of Electric Vehicle customers, but we cannot precisely measure the Range Anxiety that people feel. As a proxy, we will use the total number of public charging units. The more charging units, the less the Range Anxiety.
- Electric Vehicle jobs. If people believe that Electric Vehicles represent technological progress and an important opportunity for their economy in general and their automotive industry in particular, they may be motivated to purchase an Electric Vehicle.
- Electric Vehicle registrations. We seek to understand, explain, and predict overall Electric Vehicle registrations in general.

The three phenomena suggest a wide range of predictor variables



First EV, circa 1832



Toyota Prius 1997



Tesla Roadster 2006

Fig. 1. Evolution of Electric Vehicles.

and, more generally, a variety of variable categories. Our aims are twofold: a wide range of methods applied to a wide range of variables to obtain a highly predictive model of Electric Vehicle adoption. We set our response variables to be at the level of US state, and the predictor variables to be at the level of US county, to allow for local variation within each state. There are 3,243 counties across the United States, and we found usable data from 3,137 of them.

Methods

We sought data from many public sources to cover the categories of relevant variables. The data sources sought were high quality and credible, taken from official United States agencies: US Department of Energy, US Bureau of Labor Statistics, National Oceanic and Atmospheric Administration, Bureau of Economic Analysis, and Economic Research Service. By combining this variety of data sources, we aim to yield unique, high accuracy models, while simultaneously canceling out biases or flaws that may be attributable to individual data sources.

We checked for missing values (NULL/NA) in every variable. For a given numeric variable, we imputed for missing values with the median, except for charging units. Median is usually a reasonable replacement for missing values, assuming they are missing completely at random, because the median is resistant to outliers (whereas the mean is not). Missing values for charging units are probably not missing at random, however; they are missing because those counties simply do not yet have a significant number of Electric Vehicle drivers, and investment in national public infrastructure is currently in its early stages. Hence, we imputed those NULLs with Zero. We checked for data quality issues by checking the plots of the distributions for every variable, looking for impossible or outlying values.

The predictive models have all their response (dependent) variables in 2021 and all their predictor variables from earlier years (2010–2020). We do this to have truly predictive models, predicting over time despite the possibility of confounding influences. Table 1 shows the data sources of the response variables.

We operationalize Public Charging Units as Level 2 Charging units. We do not count Level 1 charging units as they constitute a small percentage of public charging units (approximately 4 %), equivalent to an indoor garage outlet. We do not count Level 3 charging units as they are Rapid Chargers (Direct Current), which are usable mostly by 100 % electric vehicles (not PHEVs) and often have proprietary connectors (e. g., Tesla only). Level 2 Charging units are the most universally accessible. There are 1–3 charging units per charging station. Table 2 lists all the predictor variables for which we could obtain high quality data.

We describe these variables in brief:

- A **partial ban on smoking** in certain spaces (government buildings, private businesses) indicates an emphasis on public health with a focus on air quality.
- City dwellers may have larger incomes, making them more able to consider an Electric Vehicle. However, they may have no need to do so because they have access to public transportation. Conversely,

rural dwellers may have smaller incomes and a lower cost of living, but a greater need for saving money on fuel because of longer distance commutes. **Ruralness** as a 9-point scale (from least to most rural counties) is meant to capture the impact of population density, transportation alternatives, and commuting/driving ranges.

- **The percentage of people completing college and median education** are relevant variables because more educated consumers tend to earn more and have a more positive attitude toward technological innovation (Pan et al., 2021; Pan et al., 2019).
- **Poverty Rate Percent, Gross Domestic Product, Median Household Income, and Unemployment Rate Percent** are additional economic variables that may have an effect.
- **Extant charging units** (in 2020) are relevant, because once they are installed and used, they will probably remain in place, as long as they are working.
- **Population Size** may be relevant because larger population size means greater exposure to existing Electric Vehicle owners, charging units, and shorter work commutes.
- **The Air Quality Index** represents both Public Health and the Environment.
- **Average temperature** of weather is relevant because battery performance degrades in extremely hot and extremely cold temperatures.

Although these governmental and non-governmental sources are credible data sources in themselves, we examined their data quality to assure ourselves that we were integrating accurate, complete, non-redundant information. If any variable had excessive missing values (more than 10 %) or low quality, we set it aside. If it had fewer than 10 % missing values and was reasonably clean, we cleaned it further before modeling with it.

We use the same dataset for all our hypotheses, focusing on different response variables to focus on modeling the three different issues. By doing this, we aim to gain a more holistic understanding of Electric Vehicle adoption. This can be considered slicing-and-dicing of the data from different angles, after which we will integrate our understanding of each slice, to achieve a more holistic understanding.

Results

We start with the descriptive analysis, which will be followed by the linear model of Electric Vehicle registrations and then the machine learning models of Electric Vehicle registrations. Table 3 shows the descriptive statistics for the three response variables.

Table 4 shows the descriptive statistics for the predictor variables.

We also examine a matrix plot in Fig. 4 to display the correlations among the predictor variables:

We see a wide range of correlations, from 0.0 to 0.9, with high correlations among GDP, extant Electric Vehicle charging units, and population. In any linear model we develop, we will need to show that multi-collinearity is not a problem. We also show a correlation matrix in Fig. 5, which shows the pairwise correlations broken out by median

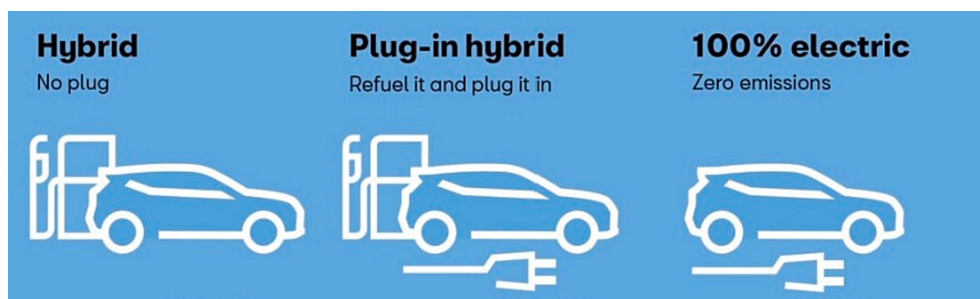


Fig. 2. Hybrid (cannot charge), Plug-in Hybrid (can charge); 100% electric (must charge). Source: www.autofutures.tv.

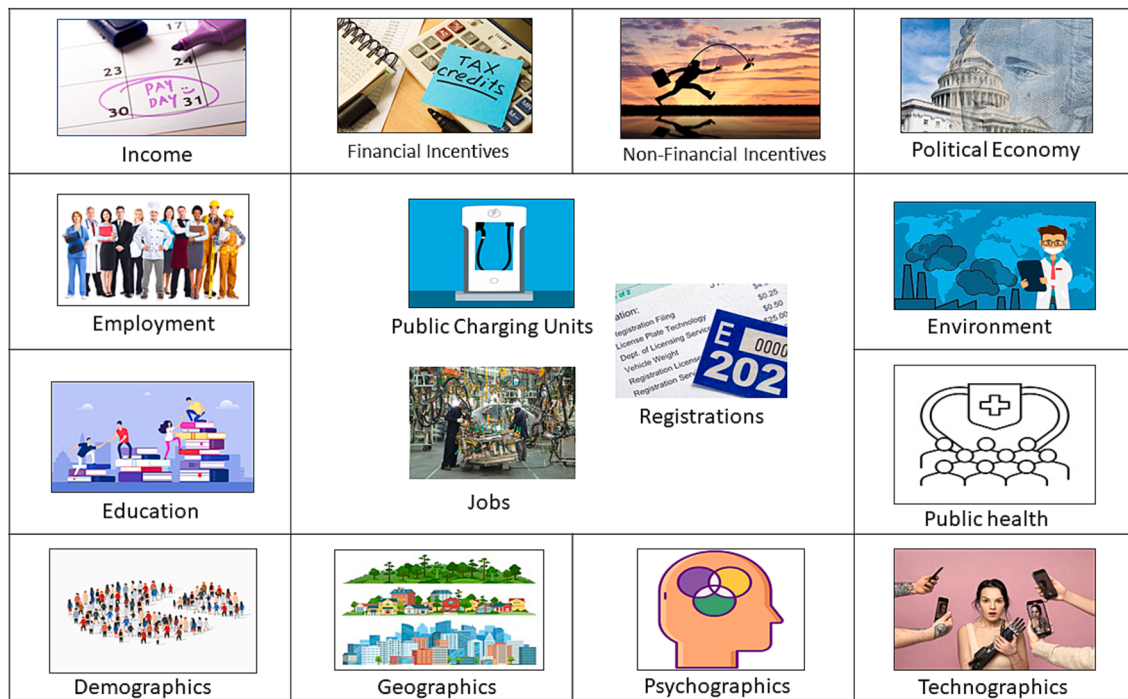


Fig. 3. Influential Factors from the Electric Vehicle literature. Sources: Employment image, Bureau of Labor Statistics; Psychographics image, Flaticon.

Table 1
Response Variables.

Response Variable	Description	Data Source and Year Retrieved	Year
Public Charging Units (Level 2 only)	Indicator of current / projected charging demand	(US Department of Energy, 2022), (Liberty Access Technologies, 2022)	2021
Electric Vehicle Jobs (hundreds)	Number of jobs in the Electric Vehicle segment of the automotive industry	US Bureau of Labor Statistics, (Indeed, 2022)	2021
Electric Vehicle Registrations	Official Electric Vehicles registered to a US state	(US Department of Energy, 2022), (Electrek, 2022)	2021

education level.

We see strong correlations among most of the predictor variables, e. g., population against GDP. Some correlations are high only among some of the median education categorical levels, e.g., temperature

against GDP only in the some-college level (in blue), which is between bachelor's degree and high school diploma as the median level of education attained. Thus, Fig. 5 suggests that education may be an important predictor.

Our modelling approach follows best practices. We randomly partition the dataset into training (80 %) and test (20 %) subsets for the ML

Table 3
Descriptive Statistics for the Response Variables, state level.

	Electric Vehicle Registrations	Electric Vehicle jobs	Public Charging Units
Min.	220	0	0
1st	2,620	8	0
Quartile			
Median	7,810	29	1
Mean	22,486	78	23.53
3rd	23,530	96	7
Quartile			
Max.	425,300	1,123	5862

Table 2
Data Sources at the US County Level.

Variable	Description	Data Source and Year Retrieved	Year	Category
Air Quality Index	Public Health / Environmental	(World Media Group, LLC, 2022)	2020	environment / public health
Average Temperature	Weather (degrees Fahrenheit)	(National Oceanic and Atmospheric Administration, 2022)	2021	environment
Completing College or University (%)	Educational Attainment	(Economic Research Service, 2022)	2015–19	education
Extant Charging Units	Demand for Electric Vehicle charging units	(US Department of Energy, 2022)	2020	technographics
Gross Domestic Product	Economy of County (units = \$1M)	(Bureau of Economic Analysis, 2022)	2019	economics / finance
Median Education	Educational Attainment	(OpenIntro, 2022)	2013–17	finance
Median Household Income	Household Income (units = \$10 K)	(Economic Research Service, 2022)	2019	finance
Partial Smoking Ban	Public Health / Environmental	(OpenIntro, 2022)	2010	public health
Population Size	Population Size (units = 10 K)	(Economic Research Service, 2022)	2020	demographics
Poverty Rate (%)	Economy of County	(Economic Research Service, 2022)	2019	economics / finance
Ruralness	Rural-Urban Continuum Codes (1 = Least Rural, 9 = Most Rural)	(US Department of Agriculture, Economic Research Service, 2022)	2013	geographics
Unemployment Rate (%)	Employment of County	(Economic Research Service, 2022)	2020	economics / finance

Table 4
Descriptive Statistics of Predictor Variables.

	Average Temp. (°F)	Gross Domestic Product (\$M)	Median Household Income (\$10 K)	Percent Completing College	Population Size (10 K)	Extant Electric Vehicle Charging Units	Percent Poverty Rate	Percent Unemployment Rate	Air Quality Index
Min.	26.6	0.000	0	0.0 %	0.006	0	2.7 %	0.0 %	21.2
1st Qt.	47.9	0.350	46.26	15.0 %	1.091	0	10.4 %	5.2 %	40.6
Median	54.3	0.930	53.36	20.0 %	2.532	1	13.4 %	6.5 %	43.9
Mean	53.9	5.894	55.72	22.0 %	10.474	13.05	14.5 %	6.7 %	43.7
3rd Qt.	60.4	2.715	62.07	26.0 %	6.720	4	17.5 %	8.0 %	46.7
Max.	70.7	726.940	151.81	78.0 %	1001.401	3113	47.7 %	22.5 %	69.2
Categorical Variable Frequencies				Median Population Education			Least	1	413
				HS diploma				2	353
				some college				3	341
				bachelor's				4	206
					Ruralness			5	213
								6	571
								7	420
				Smoking Ban				8	216
				none				9	404
				partial			Most		



Fig. 4. Correlations Among Predictor Variables.

models. We transform variables to be more normally distributed, to be in greater compliance with the statistical assumptions of linear models.

We start with multiple regression as a method for obtaining baseline accurate predictive models, all of which use independent variables from years prior to that of the dependent variable. The functional form of a regression can be described as in equation (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \tag{1}$$

We compare a random effects model to a fixed effects model to select the more appropriate one. After showing each regression model, and showing the diagnostics, we will interpret the beta coefficients. We will then try to obtain more accurate machine learning models using more advanced machine learning methods, to predict the same three response

variables. Those models will likely be more accurate, because they can capture non-linearities.

Linear regression results

We first develop linear regression models for our three Electric Vehicle phenomena of interest: Public Charging Units, Electric Vehicle Jobs, and Electric Vehicle Registrations. We estimate both a fixed effects model, indexed by US state, and a random effects model.

Public Charging Units

To determine whether a fixed effects or random effects model is more appropriate, the Hausmann test was performed. The F-statistic on the



Fig. 5. Correlations between key predictor variables, broken out by median education level.

chi-squared test of the fixed effects model coefficients and random effects model coefficients yielded a p-value > 0.05 . The null hypothesis was therefore not rejected, thereby determining the random effects model as the more appropriate model. Table 5 shows the output from the best multiple regression, with eight of the predictor variables being statistically significant.¹ The output is sorted by t-value.

The variance explained (adj. R^2) of the model is 62.44 %. A check for collinearity indicates no problems, with all Variance Inflation Factor scores < 5 . Examining all the coefficient estimates, we consider them in both sign and magnitude.

Solar power for generating electricity has the highest statistical significance and magnitude. An extra unit of solar generation means an additional 927 charging units. That suggests that populations investing in solar energy are correlated positively with Electric Vehicle charging

Table 5
Regression Results for Public Charging Units in 2021.

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	48026.362	2799.100	17.158	***
Solar_power_e_gen_2019	927.362	20.742	44.709	***
Extant_charging_units_2020	7.027	0.791	8.879	***
Household_Income_2019	188.272	81.534	2.309	*
Coll_grad_rate_2015-2019	-2500.501	988.641	-2.529	*
Poverty_rate_2019	-4703.377	1688.496	-2.786	**
Air_Quality_Index_2020	-63.470	14.281	-4.444	***
Partial_smoking_ban_2010	-1164.517	170.281	-6.839	***
HS_grad_rate_2015-2019	-49558.111	2754.677	-17.991	***

Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Residual standard error: 3108 on 2184 df.

Multiple R-squared: 0.6258, Adj. R-squared: 0.6244.

F-stat: 456.5 on 8 and 2184 DF, p-value: $< 2.2e-16$.

¹ Although median education level was helpful in the descriptive statistics, it was not significant in any regression.

units. There is a smaller but positive coefficient (7.027) on the number of extant charging units, indicating approximately seven additional chargers in 2021 for every charger in 2020. Household income has a positive coefficient (188.272), indicating that for every \$10,000 of household income in 2019, there are approximately 188 more charging units two years later, in 2021. Similarly, lower poverty is positively associated with Electric Vehicles. For every percentage decrease in poverty, the charging units increase by 4,703.

College graduation rate in 2015–2019 has a small negative effect, indicating that a one percent increase in graduation rate means a 2,500 decrease in the (public) Electric Vehicle Charging units in 2021. Consistent with that finding is the High School graduate Rate (2015–2019), indicating a decrease of 49,558 charging units for every percent increase in the high school graduation rate. It seems that greater education corresponds with greater income and likelihood of private charging (at home or at work). The Air Quality Index (higher is worse air quality) in 2020 is highly significant; every additional unit means 63 fewer charging units in 2021. This suggests that people who live where Air Quality is worse are less able or willing to purchase an Electric Vehicle. A Partial Smoking Ban is negatively associated with Electric Vehicle charging units, indicating that the banning of smoking in certain areas is associated with fewer public charging units. This counter-intuitive result may indicate a community where public health regulations curtail smoking, but people own their own home, where they charge (and smoke) privately, or take public transportation rather than drive. Overall, the model explains most of the variance in public charging units, and that the strongest predictors are solar power, high school education rate, household income, and air quality index.

Electric Vehicle jobs

To determine whether a fixed effects or random effects model is more appropriate, the Hausmann test was performed. The F-statistic on the chi-squared test of the fixed effects model coefficients and random effects model coefficients yielded a p-value > 0.05 . The null hypothesis was therefore not rejected, thereby determining the random effects model as the more appropriate model. Table 6 shows the output from the

best multiple regression, with seven of the predictor variables being statistically significant. The output is sorted by t-value.

The variance explained (adj. R^2) of the model is 51.52 %. A check for collinearity indicates no problems, with all Variance Inflation Factor scores < 5. Examining all the coefficient estimates, we consider them in both sign and magnitude.

The predictor with the highest statistical significance is solar power generation. An extra unit of solar generation means an additional 2,770 Electric Vehicle jobs. That suggests that solar power is positively associated with Electric Vehicle employment, probably because the same population that supports solar power (rather than fossil fuels) also supports Electric Vehicle jobs. For example, the population could support the retraining of fossil fuel workers to work in the solar or Electric Vehicle sector. Extant charging units in 2020 are associated with Electric Vehicle jobs; every unit increase is associated with nearly 23 additional Electric Vehicle jobs. Household income is not statistically significant, but poverty rate (2019) is negatively associated; every percent decrease in poverty rate is associated with nearly 18,702 additional Electric Vehicle jobs. The Air Quality Index is associated negatively with Electric Vehicle jobs. Every unit decrease of AQI (improved air quality) is associated with 180 more Electric Vehicle jobs. A partial smoking ban in 2010 is associated with 4,583 fewer Electric Vehicle jobs in 2021.

The negative coefficients on high school graduation rate and College graduation rate are associated with fewer Electric Vehicle jobs. This could be that people that are more educated tend to take jobs outside of the automotive sector, or that they tend to be recent graduates, who live in apartments, cannot afford an Electric Vehicle, or are waiting for the Electric Vehicle industry to mature. Overall, this model explains most of the variance in Electric Vehicle jobs, and that the strongest predictors are solar power and high school graduation rate.

Electric Vehicle registrations

To determine whether a fixed effects or random effects model is more appropriate, the Hausmann test was performed. The F-statistic on the chi-squared test of the fixed effects model coefficients and random effects model coefficients yielded a p-value > 0.05. The null hypothesis was therefore not rejected, thereby determining the random effects model as the more appropriate model. Table 7 shows the output from the best multiple regression, with eight of the predictor variables being statistically significant. The output is sorted by t-value.

The adjusted variance explained is 62.09 % with all the predictors statistically significant. A check for collinearity indicates no problems, with all Variance Inflation Factor scores < 5. Examining all the coefficient estimates, we consider them in both sign and magnitude.

The predictor with the highest statistical significance is solar power generation of electricity. An extra unit of solar generation means an additional 10,687 Electric Vehicle registrations. That suggests that populations that support solar energy also purchase Electric Vehicles. It may again indicate that solar power is partly being used to generate the

Table 7

Regression Results for Electric Vehicle Registrations in 2021.

	Estimate	Std. Error	T value	Pr(> t)
(Intercept)	564189.060	32461.520	17.380	***
Solar_power_e_gen_2019	10686.897	240.551	44.427	***
Extant_charging_units_2020	82.279	9.178	8.965	***
Household_Income_2019	2589.605	945.554	2.739	**
Poverty_rate_2019	-53662.278	19581.700	-2.740	**
Coll_grad_rate_2015-2019	-39542.521	11465.397	-3.449	***
Partial_smoking_ban_2010	-12359.404	1974.766	-6.259	***
Air_Quality_Index_2020	-1044.238	165.613	-6.305	***
HS_grad_rate_2015-2019	-575025.797	31946.343	-18.000	***

Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Residual standard error: 36,050 on 2184 df.

Multiple R-squared: 0.6223, Adj. R-squared: 0.6209.

F-stat: 449.8 on 8 and 2184 DF, p-value: < 2.2e-16.

electricity needed for the Electric Vehicles, or it could be an indirect effect. Extant charging units in 2020 have a positive effect; every extra unit is associated with 82 additional Electric Vehicle registrations. Household income and poverty rate have a similar effect. Every \$10,000 increase in median household income corresponds to an additional 2,590 Electric Vehicle registrations and every percent decrease in poverty corresponds to an additional 53,662 Electric Vehicle registrations.

The AQI with a negative coefficient suggests that every additional unit (worse air quality) is associated with 1,044 fewer charging units. This suggests that people who live or work with better air quality have more Electric Vehicles. The Electric Vehicles may be improving the air quality directly through fewer tailpipe emissions, or it may reflect a population that is more able and willing to purchase an Electric Vehicle. A partial smoking ban has a negative coefficient, indicating that the banning of smoking in higher population density areas (e.g., cities) is associated with 12,359 fewer registrations. This may suggest a community where public health regulations curtail smoking, but people take public transportation rather than drive Electric Vehicles. The strongly negative coefficient on high school graduation rate is associated with fewer registrations. A 1 % increase in the high school graduation rate corresponds to 575,026 fewer Electric Vehicle registrations. The college graduation rate effect is similar, but much smaller. A 1 % increase in the college graduation rate corresponds to 39,543 fewer Electric Vehicle registrations. These findings suggest that recent graduates live in apartments, cannot afford an Electric Vehicle, or are waiting for the Electric Vehicle industry to mature. Overall, this model explains most of the variance in Electric Vehicle registrations, and that the strongest predictors are solar power, high school graduation rate, and extant charging units.

Summary of linear models

Overall, the three regressions give us reasonably strong predictive models. The adj. R^2 ranges between 51.52 % and 62.44 %. Most of the predictor variables are significant in most of the three models, and although the beta coefficient magnitudes vary depending on model, the results are consistent within and across the models. In all three models, the collinearity was checked and determined to be low. However, the models may have failed to capture interaction effects or other non-linearities. Rather than create interaction terms and higher order polynomial terms for all combinations of the eight predictors, we opted to use well-established machine learning methods to capture non-linear relationships. Since these methods quickly become a "black box," difficult to explain and interpret, we also examine SHAP plots (Fidel et al., 2020; Kaur et al., 2020) to understand them in detail.

Table 6

Regression Results for Electric Vehicle Jobs in 2021.

	Estimate	Std. Error	T value	Pr(> t)
(Intercept)	1331.109	103.181	12.901	***
Solar_power_e_gen_2019	27.704	0.765	36.233	***
Extant_charging_units_2020	0.226	0.029	7.738	***
Household_Income_2019	3.949	3.006	1.314	
Coll_grad_rate_2015-2019	-104.174	36.443	-2.859	**
Poverty_rate_2019	-187.026	62.242	-3.005	**
Air_Quality_Index_2020	-1.799	0.526	-3.417	***
Partial_smoking_ban_2010	-45.830	6.277	-7.301	***
HS_grad_rate_2015-2019	-1334.827	101.543	-13.145	***

Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 114.6 on 2184 df

Multiple R-squared: 0.517, Adj. R-squared: 0.5152

F-stat: 292.2 on 8 and 2184 DF, p-value: < 2.2e-16

Machine learning (ML) models

ML methods are effective at capturing linear and non-linear relationships (Adadi & Berrada, 2018; Kavadi et al., 2020). We applied a variety of machine learning methods using R and various packages. Specifically, the algorithms evaluated in this paper are Generalized Additive Models, Neural Networks, and Gradient Boosted Trees, because they are the type of models that tend to achieve high accuracy in the literature and in data science competitions (Ganjisaffar et al., 2011).

- Generalized Additive Model (GAM): a generalization of linear regression; allows for nonlinear relationships and coefficient regularization. Instead of assuming a linear relationship, a GAM allows for more flexible, non-linear relationships between the variables. The equation can be described as in equation (2):

$$Y = f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) + \varepsilon \quad (2)$$

Although overfitting can occur, regularization and cross-validation help to minimize it. (R package mgcv).

- Neural Network: a generalization of linear regression with one or more hidden layer(s) of nodes between the input and output nodes in which the hidden layer represents a set of non-linear transformations of the input variables. Neural networks allow for different activation functions. An arbitrarily complex model can be fitted because of the hidden layers and non-linear activation functions. It may result in overfitting.

$$Y = f(\beta_0 + \beta_1 h_1(X_1) + \beta_2 h_2(X_2) + \dots + \beta_p h_p(X_p)) + \varepsilon \quad (3)$$

In equation (3), described above, $h_1(X_1)$, $h_2(X_2)$, ..., $h_p(X_p)$ are the outputs of the neurons in the hidden layer, which are functions of the input variables X_1 , X_2 , ..., X_p . The coefficients β_0 , β_1 , β_2 , ..., β_p are the weights and bias terms of the model, which are learned during the training process. The function f is the activation function applied to the sum of the weighted inputs and bias terms in the hidden layer (R package neuralnet).

Gradient Boosted Tree (GBT): Finally, rather than another method with a closed-form mathematical representation, we choose a low bias method consisting of a sequence of regression trees in which the residuals of the tree are modeled by another tree. The gradient boosting algorithm iteratively adds decision trees, creating a sequential ensemble (Bolón-Canedo & Alonso-Betanzos, 2019) by using gradient descent search to minimize loss of a differentiable loss function. Parameters include learning rate, minimum or maximum number of trees, regularization parameter, and pruning parameter. The local minimum found is likely to be the global minimum if many trees are fitted with random starting points (R package XGBoost).

All these machine learning methods were selected for their known strengths in minimizing Errors of Bias and/or Errors of Variance (Geman et al., 1992), which helps them to generalize, i.e., fit the data well on holdout, test data without overfitting (Ng, 1997).

We describe the results of the ML models for all three models. For accuracy metrics, we use, in addition to variance explained for the training data, two standard measures of predictive accuracy on the test data: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (Hyndman & Koehler, 2006; Willmott & Matsuura, 2005), and compare them to those of multiple regression. To explain the most accurate models, we show SHAP plots.

Public charging units

In the multiple regression model having 62.4 % variance explained, the key predictors are solar power, education, household income, and air quality. All three ML models perform better, as can be seen in Table 8, using 80/20 split on Train/Test and 5-fold cross-validation. Gradient Boosted Tree performs the best, with an adj. R^2 of 98.8 % on

Table 8

Accuracy metrics for Public Charging Units.

method	Train			Test	
	adj. R^2	RMSE	MAE	RMSE	MAE
Gradient Boosted Tree	0.988	519.23	287.69	633.50	305.74
Generalized Additive Model	0.844	1846.47	551.65	1461.37	531.29
Neural Network	0.678	2650.74	1135.52	3020.92	1277.28
Multiple Regression	0.624	2739.44	1731.24	3187.17	1978.56

the training data and accuracy metrics far better than the other models on the test data. The main parameters for the GBT are gamma loss function, learning rate = 0.05, early stopping allowed after 200 iterations, and no maximum tree depth.

The lift chart for the GBT can be seen in Fig. 6, showing consistently accurate predictions.

The GBT is an extremely accurate model. The plot in Fig. 7 shows that the top two predictors have far larger impact than the other predictors, although all eight do contribute to the accuracy. This illustrates that solar power generation and high school graduation rate have far greater importance (and impact) than air quality index, extant charging units, household income, poverty rate, college graduation rate, and partial smoking ban.

The SHAP plots show further that the influence is non-monotonic, i.e., that low, middle, and high range values of most of the predictors do not trace a simple, linear relationship.

Solar power and high school graduation rate have the greatest importance. The solar relationship is positive, as can be seen in Figs. 8 and 9. The low values of solar (yellow) correspond to lower public charging units and high values (purple) to the right of the zero point. This indicates that solar power and charging units are positively related, but the SHAP plot shows a sharply positive slope followed by a negative slope and then a steady increasingly positive slope. The relationship between high school graduation and public charging units is also non-monotonic. It is a negative relationship at the extremes but positive relationship in the middle. If we look at the middle of each chart, ignoring the extremes completely, we see a moderately positive slope in both.

Electric vehicle jobs

The multiple regression model has variance explained of 51.5 %. Two of the machine learning models perform better, as can be seen in Table 9, using 80/20 split on Train/Test and 5-fold cross-validation. Gradient Boosted Tree performs the best, with an adj. R^2 of 89.7 % on the training data and accuracy metrics far better than the other models on the test data. The main parameters for the GBT are least-squares loss function, learning rate = 0.05, no early stopping, and no maximum tree depth. The lift chart for the GBT can be seen in Fig. 10, showing consistently accurate predictions.

Fig. 11 shows that the top predictors, high school graduation rate and solar power generation, have a far larger impact than the other predictors, although the other six contribute to the accuracy.

Fig. 12 shows a big cluster of low yellow dots on high school graduation, which indicates a very positive impact on Electric Vehicle jobs. The rest of that plot line shows a mixture of yellow and purple dots. All other factors have much smaller impact.

The GBT is an extremely accurate model of Electric Vehicle jobs. The SHAP plots in Fig. 13 show further that the influence is non-monotonic, i.e., that low, middle, and high range values of the predictors do not trace a simple, linear relationship.

Electric Vehicle registrations

Electric Vehicle registrations represent Electric Vehicle sales regardless of the consumer's motivations, reasons, or concerns. All the ML models perform better than Multiple Regression, as can be seen in

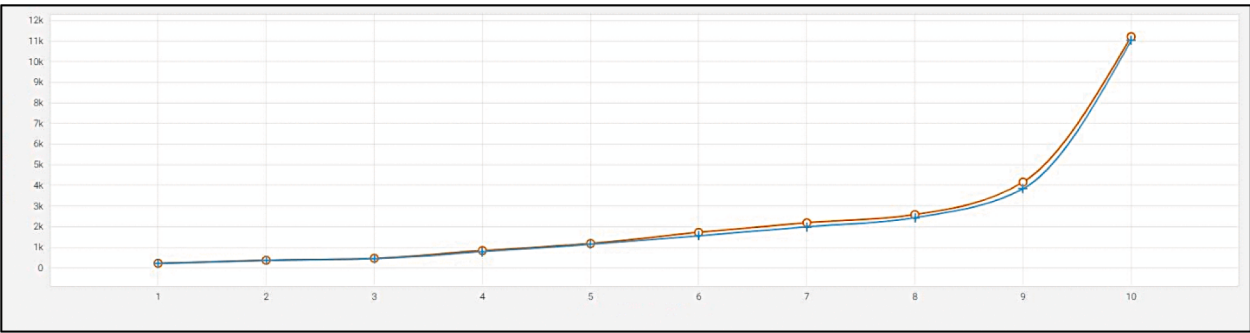


Fig. 6. Lift Chart: Average Public charging units ~ Predicted Value Bins; Actual Value, Predicted Value.

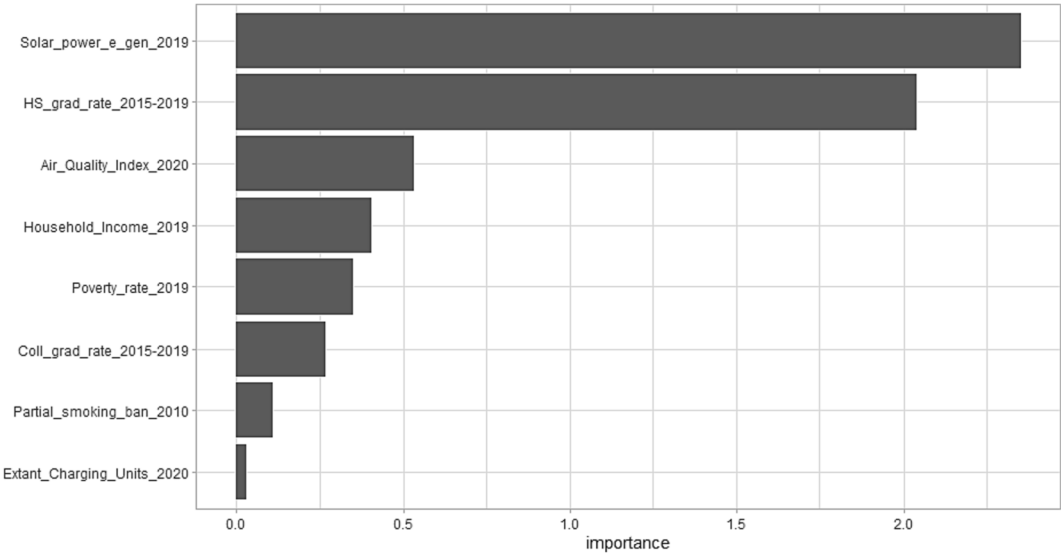


Fig. 7. Variable Importance Plot.

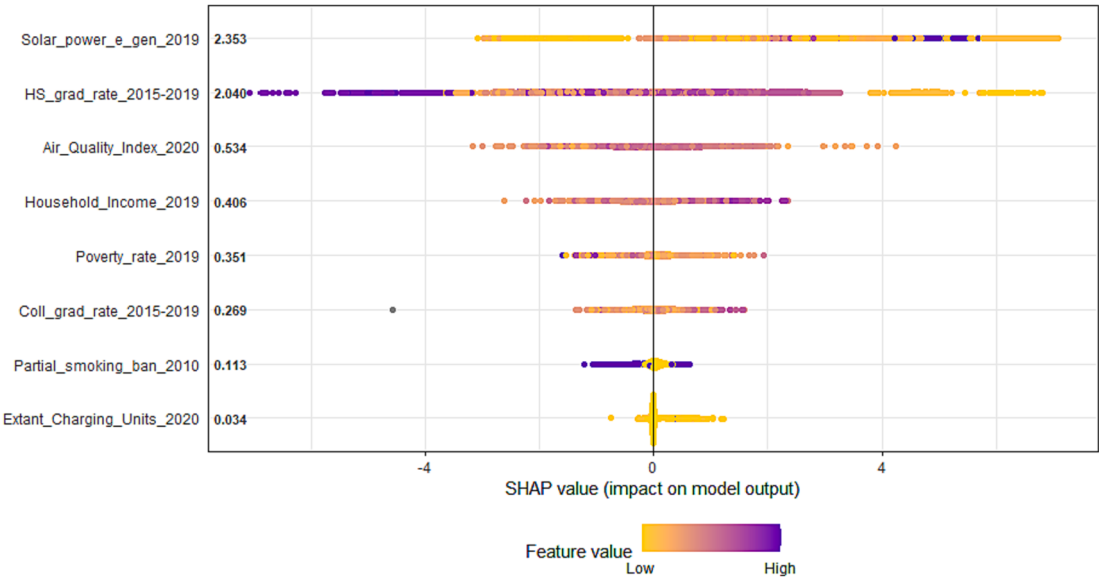


Fig. 8. SHAP plot for Public Charging Units, All Variables.

Table 10, using 80/20 split on Train/Test and 5-fold cross-validation. Gradient Boosted Tree performs the best, with an adj. R2 of 98.4 % on the training data and accuracy metrics far better than the other models

on the test data. The main parameters for the GBT are gamma loss function, learning rate = 0.05, early stopping allowed after 200 iterations, and no maximum tree depth.

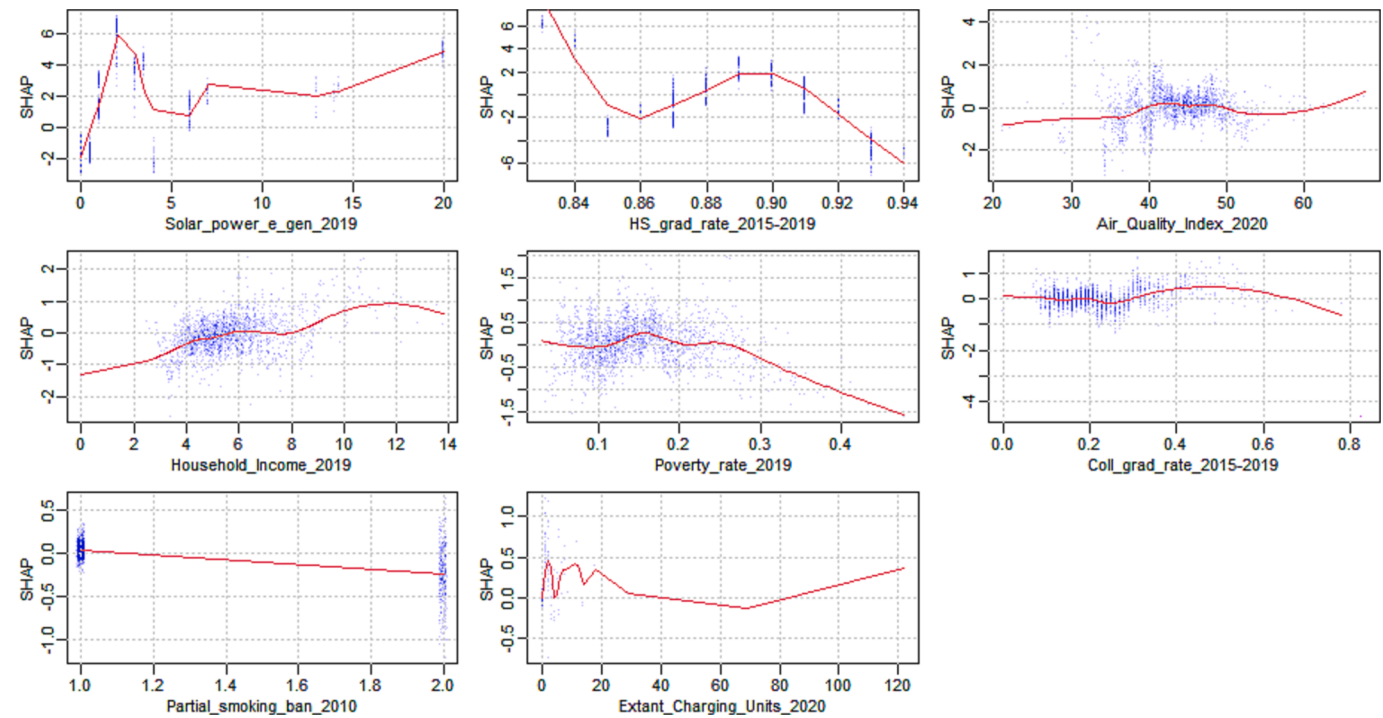


Fig. 9. SHAP plots for Public Charging Units to show the influence of each predictor variable.

Table 9
Accuracy metrics for Electric Vehicle Jobs.

method	Train			Test	
	adj. R^2	RMSE	MAE	RMSE	MAE
Gradient Boosted Tree	0.897	48.99	20.14	52.28	20.95
Generalized Additive Model	0.870	55.04	26.80	54.01	26.98
Multiple Regression	0.515	103.97	61.89	115.00	67.56
Neural Network	0.271	130.18	68.34	137.39	71.42

The GBT is an extremely accurate model of Electric Vehicle registrations. Figs. 14-15 not only explain the GBT, but they show that the top two predictors, solar power generation and high school graduation rate, have a far larger impact than do the other predictors. Figs. 16-17 show further that the influence is non-monotonic, i.e., that low, middle, and high range values of all the predictors have non-linear effects.

These plots show a similar story regarding solar power, a positive correlation. More electricity generation from solar corresponds to greater Electric Vehicle registrations. It could be direct, in that homeowners with solar roofs believe in renewable energy more generally, or it could be indirect, that populations who support solar power, e.g., community solar or solar farms, are more likely to install public

charging units.

The plots show an interesting relationship between education and electric vehicle registrations. It seems that education is an important if non-linear predictor. It could be that some less educated people are not interested, cannot afford an Electric Vehicle, or do not have access to private charging, whereas other less educated people are interested in an Electric Vehicle but bidding their time, waiting for more affordable models or more ubiquitous public charging. For the more educated people, the lower area of the stripe may indicate those who purchase one eagerly for environmental reasons. The higher area of the stripe may indicate more educated people who are interested but not ready for some reason. Overall, if one ignores the extremes, there is a positive relationship between education and Electric Vehicle registrations.

Discussion

To understand and model Electric Vehicle adoption, we aggregated data from multiple sources in multiple areas, including finance, health, environment, and education, and developed both linear and nonlinear models. Ultimately, we combined a variety of factors to arrive at three models to predict accurately the response variables: public charging units, jobs, and registrations. This set of highly accurate models is a novel contribution to the Electric Vehicle research literature.

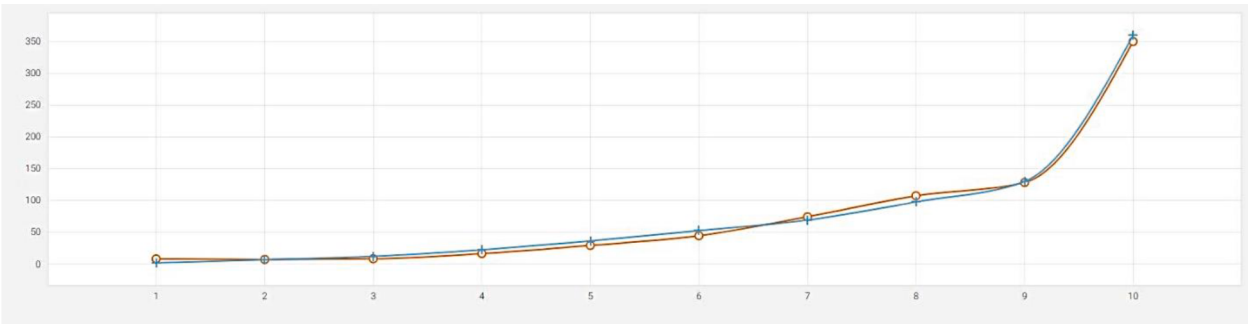


Fig. 10. Lift Chart: Average Electric Vehicle Jobs ~ Predicted Value Bins; Actual Value, Predicted Value.

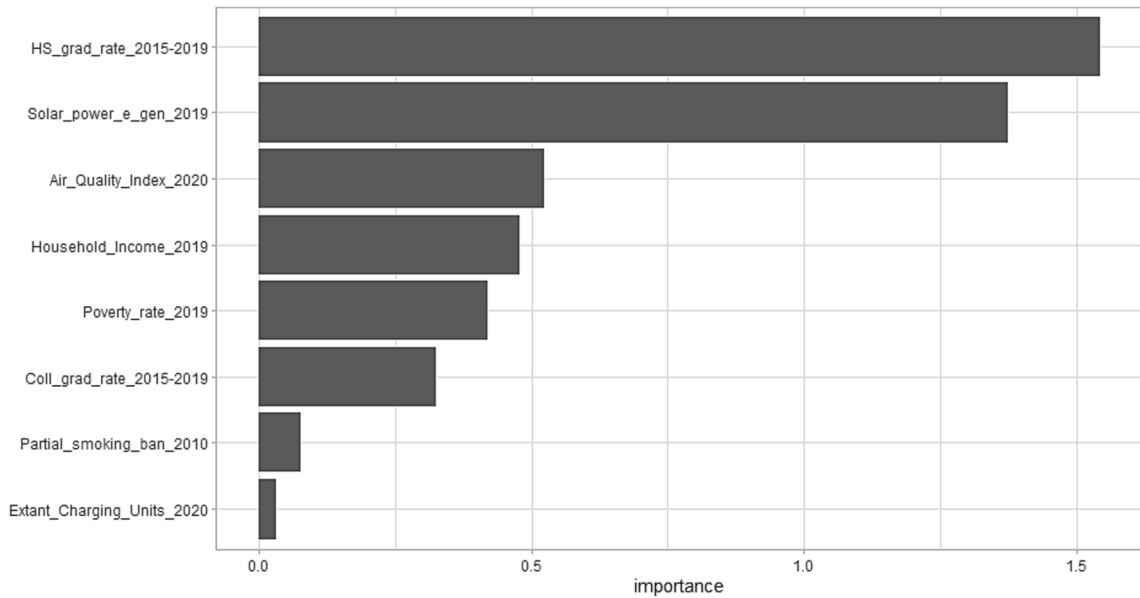


Fig. 11. Variable Importance Plot, Electric Vehicle Jobs.

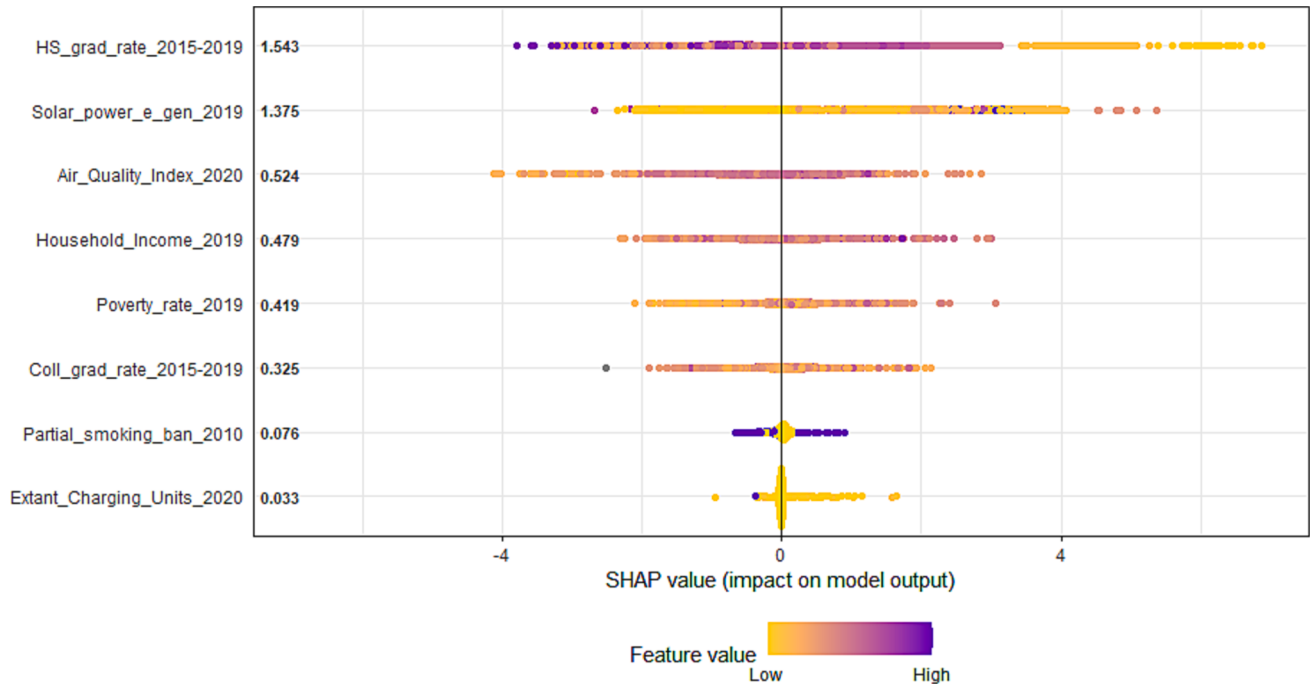


Fig. 12. SHAP plot for Electric Vehicle Jobs, All Variables.

All the best performing predictive models are Gradient Boosted Trees, with high variance explained (adjusted $R^2 = 0.988$, 0.897 , and 0.984 on the training data partitions) and the highest predictive accuracy of the machine learning on the test data partitions. Gradient Boosted Trees are ensemble models, which are often considered black-boxes, difficult to interpret and understand. We opened the “black-boxes,” by using variable importance plots and SHAP plots to enhance explainability and interpretability (Adadi & Berrada, 2018; Gilpin et al., 2018). They showed that the top two predictors were high school graduation rates and solar power generation of electricity, with air quality index in third place. The other predictor variables – partial smoking ban, college graduation rate, poverty rate, household income, and extant charging units – had a significant, but smaller impact.

Our original contributions in this paper include 1) integration of multiple publicly available data sources of Electric Vehicle variables, 2) highly accurate predictive models, and 3) comparison and contrast of multiple regression models with machine learning models. The Gradient Boosted Trees enhance our understanding gained from the Multiple Regression models. They also add nuance by capturing non-linearities. They highlight the influence of high school graduation rates, solar power generation of electricity, and air quality index as the most important predictors in predicting all three phenomena: public charging units, electric vehicle jobs, and electric vehicle registrations. The non-linearities among those predictors — in which high and low ranges of a given predictor had a sign opposite to that of the middle range — deserve future investigation.

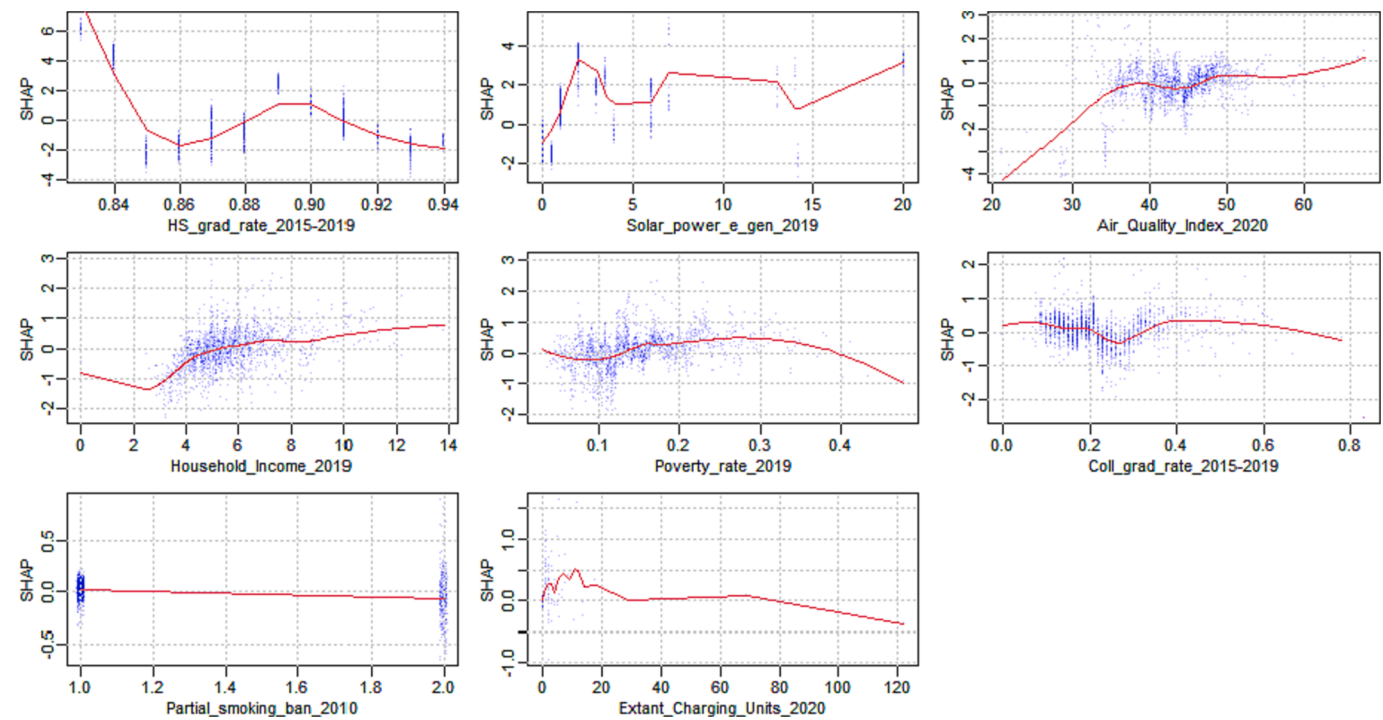


Fig. 13. SHAP plots for Electric Vehicle Jobs to show the influence of each predictor variable.

Table 10
Accuracy Metrics for Electric Vehicle Registrations.

method	Train			Test	
	adj. R ²	RMSE	MAE	RMSE	MAE
Gradient Boosted Tree	0.984	6905.96	2671.20	7347.91	2703.43
Generalized Additive Model	0.663	31313.00	7335.16	20391.00	6181.10
Neural Network	0.676	30668.00	12972.00	35356.00	14615.00
Multiple Regression	0.621	31796.00	19502.00	36787.00	22205.00

Implications for Research, Practice, and policymakers

Our models explain the vast majority of the variance on the training partition and perform very strongly on the holdout, test data, with minimal deterioration when comparing accuracy between the train and test partitions. The accuracy could be improved further, however, particularly in the Electric Vehicle jobs model. We could attempt to capture additional accuracy with interactions between pairs of features or higher order polynomial versions of them. Beyond feature engineering, we could collect data on new, additional features.

Our strongly predictive models suggest that the small but significant impacts from a partial smoking ban, college graduation rate, poverty rate, household income, and extant charging units could be of interest to automakers, car dealers, and policy makers. Financial and non-financial incentives may need to vary by geography, demographics, and other variables. Our models have implications for where Electric Vehicles can be sold, specifically in areas currently using or receptive to using solar power for generating electricity. The strong relationship could be as a direct influence, i.e., solar power to generate cheap electricity for the Electric Vehicles, or it could indicate an indirect influence: an underlying attitude for technological innovation and against fossil fuels. These models can help determine the optimal location for new Electric Vehicle charging units.

There are several potential benefits of Electric Vehicles that could be particularly appealing to people with lower levels of education attainment. First, Electric Vehicles are generally more affordable to operate than gasoline-powered cars because they are cheaper to fuel and require less maintenance. They may also be eligible for means-tested government incentives and tax credits to make the purchase price more affordable. An Online Vehicle Energy Cost Calculator has been developed and tested in an effort to educate and persuade consumers who may have difficulty assessing of the tradeoffs involved (Sanguinetti

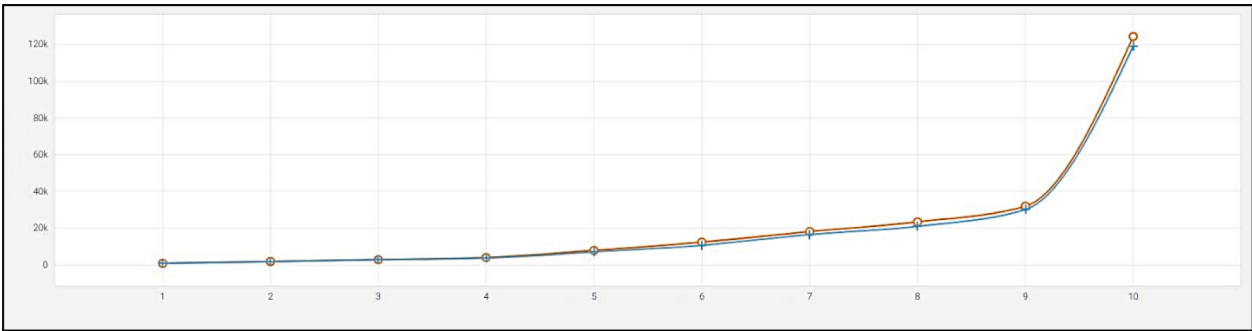


Fig. 14. Lift Chart: Average Electric Vehicle Registrations ~ Predicted Value Bins; Actual Value, Predicted Value.

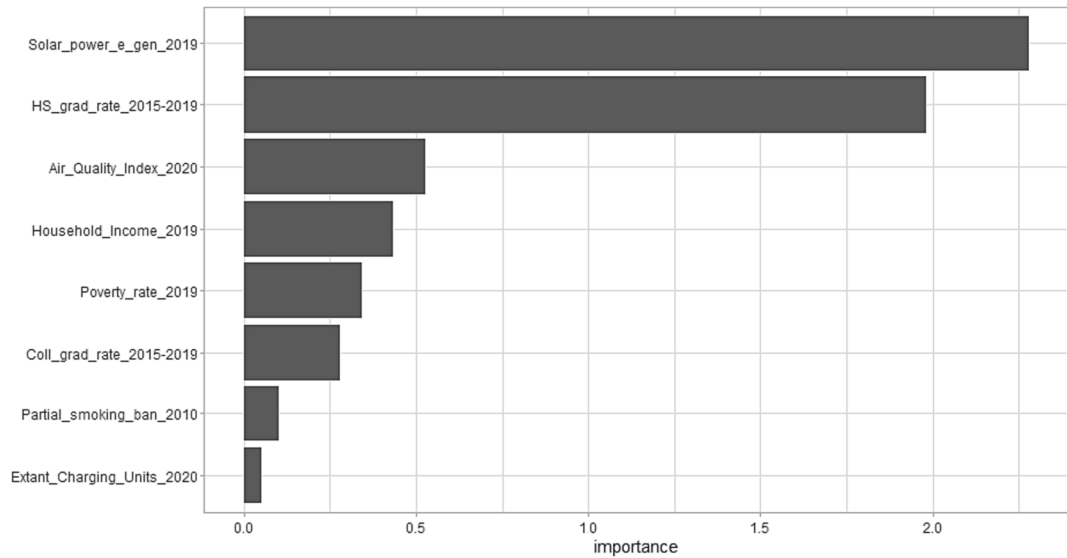


Fig. 15. Variable Importance Plot.

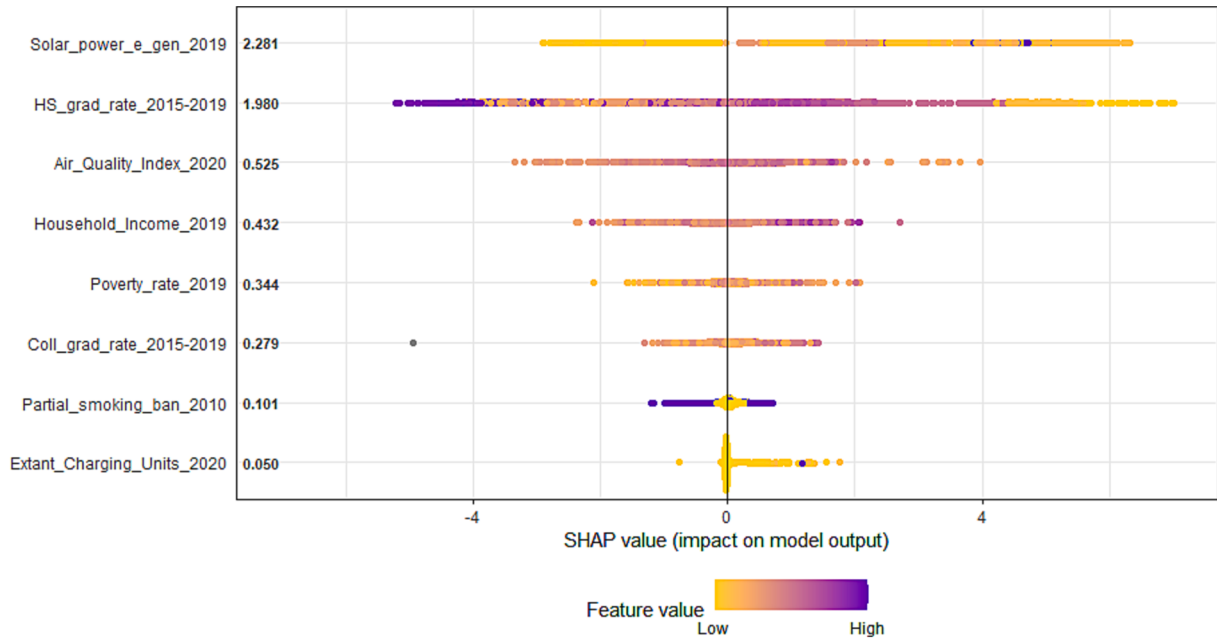


Fig. 16. SHAP plot for Electric Vehicle Registrations, All Variables.

et al., 2017). Finally, potential customers of lower educational attainment could be persuaded to choose a career path through vocational-technical schools in order to work in the Electric Vehicle industry.

Electric Vehicle automakers and car dealers know that their profits must come from purchase price, not maintenance, since Electric Vehicles require less maintenance than ICE vehicles. Dealers are pricing the Electric Vehicles accordingly, but how can marketers use that information? Do they target the owners of solar panels? Do they market to people who live near public charging units or those who live in towns/cities dedicated to solar power generation? This research suggests that they should do so. In addition, this research suggests that marketing and advertising budgets could focus on demographics of the customer (household income, college graduation rate) in addition to aspects of the county (partial smoking ban, charging units, extant charging units).

Other research has shown some symbiotic effects between Electric Vehicle owners and their communities.

- Parking lot owners can install solar energy canopies to protect and charge cars parked beneath them (Figueiredo et al., 2017) or have solar-powered battery backup for Electric Vehicle charging stations (Zhao & Burke, 2014), thus enjoying multiple levels of renewable energy.
- Used car batteries at the end of their lifespan have been re-deployed on solar parking lots roofs to enable the capturing of new solar power (Assunção et al., 2016).
- Owners of Electric Vehicles that have more electricity than they need can sell their excess energy back to the grid (Vehicle to Grid), particularly during peak energy usage times, thus engaging in electricity arbitrage (Richardson, 2013).
- Neighborhood microgrids, energy trading, and vehicle sharing are possible, changing incentives across the country for people with Electric Vehicles, house batteries, and solar panels (Meisel & Merfeld, 2018).

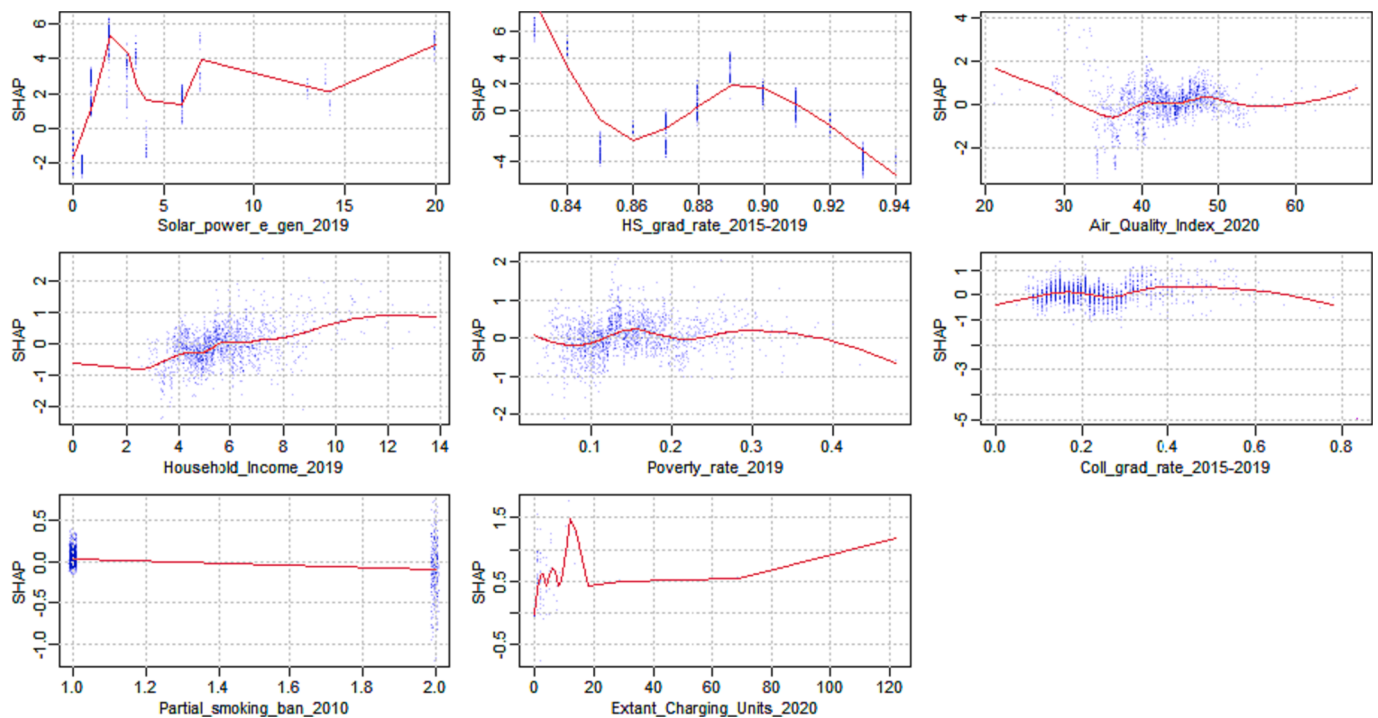


Fig. 17. SHAP plots for Electric Vehicle Registrations to show the influence of each predictor variable.

These trends all suggest new configurations of renewable energy to consider, benefiting Electric Vehicle owners and their communities.

Limitations and future directions

We developed strongly predictive models of a rapidly changing phenomenon, electric vehicles, based on a diverse set of data sources and modeling methods. The models helped us to identify the core factors influencing electric vehicle adoption: solar generation of electricity, high school graduation rates, and air quality. This is a core contribution, filling a gap in our understanding of EV adoption as it relates to buyer characteristics and their environment. This contribution could help policymakers and marketers alike. Nevertheless, this research has some limitations. We have no data indicating car ownership rates, home charging, workplace charging per county, or political leanings, i.e., liberal vs. conservative counties. In addition, the lack of statistical significance of ruralness could be attributed to the confounding of public transportation with other factors, e.g., regional or state variation.

Future research could set aside California – the Electric Vehicle leader – as well as parts of the industrial Midwest (Michigan, Illinois, Indiana, Ohio, Wisconsin), industrial Mid-Atlantic (New York, Pennsylvania), and industrial Appalachia (West Virginia, Kentucky), where most of the ICE car manufacturing takes place. The data was gathered from 3,137 of the 3,243 counties of the United States because not all counties have collected and reported data. We excluded the US territories because complete data was not available for American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, or the U.S. Virgin Islands. We may have missed some nuances in our models owing to the fast-evolving nature of the electric vehicle revolution. These limitations suggest additional variables that could be added in the future after Electric Vehicles have achieved widespread adoption, nationwide infrastructure, and market equilibrium.

CRedit authorship contribution statement

Arnold Kamis: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Data curation,

Conceptualization. **Preethi Susan Abraham:** Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. : This research received no external funding. The APC was funded by Brandeis University, International Business School.

Data availability

Data will be made available on request.

Acknowledgements

We thank Cynthia Kipkorir and Venkata Sai Nikhil Teja Nama for their assistance during this project. We also thank anonymous reviewers at the International Business School, Brandeis University, for their helpful comments.

Funding

This research received no external funding. The APC was funded by Brandeis University, International Business School.

References

- Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>.
- Afshar, S., Macedo, P., Mohamed, F., Disfani, V., 2021. Mobile charging stations for electric vehicles — A review. *Renew. Sustain. Energy Rev.* 152, 111654 <https://doi.org/10.1016/j.rser.2021.111654>.
- Andre, M., Sartelet, K., Moukhtar, S., Andre, J.M., Redaelli, M., 2020. Diesel, petrol or electric vehicles: What choices to improve urban air quality in the Ile-de-France region? A simulation platform and case study. *Atmos. Environ.* 241, 117752 <https://doi.org/10.1016/j.atmosenv.2020.117752>.
- Assunção, A., Moura, P.S., de Almeida, A.T., 2016. Technical and economic assessment of the secondary use of repurposed electric vehicle batteries in the residential sector to

- support solar energy. *Appl. Energy* 181, 120–131. <https://doi.org/10.1016/j.apenergy.2016.08.056>.
- Bolón-Canedo, V., Alonso-Betanzos, A., 2019. Ensembles for feature selection: A review and future trends. *Information Fusion* 52, 1–12. <https://doi.org/10.1016/j.inffus.2018.11.008>.
- Bureau of Economic Analysis (2022). <http://bea.gov>.
- Christidis, P., Focas, C., 2019. Factors affecting the uptake of hybrid and electric vehicles in the European Union. *Energies* 12, 3414. <https://doi.org/10.3390/en12183414>.
- Cunha, Á., Brito, F.P., Martins, J., Rodrigues, N., Monteiro, V., Afonso, J.L., Ferreira, P., 2016. Assessment of the use of vanadium redox flow batteries for energy storage and fast charging of electric vehicles in gas stations. *Energy* 115, 1478–1494. <https://doi.org/10.1016/j.energy.2016.02.118>.
- Economic Research Service. (2022). <http://data.ers.usda.gov>.
- US Department of Agriculture, Economic Research Service. (2022). <http://www.ers.usda.gov>.
- Electrek. (2022). electrek.co.
- Fidel, G., Bitton, R., & Shabtai, A. (2020, 19–24 July 2020). *When Explainability Meets Adversarial Learning: Detecting Adversarial Examples using SHAP Signatures*. International Joint Conference on Neural Networks.
- Figueiredo, R., Nunes, P., Brito, M.C., 2017. The feasibility of solar parking lots for electric vehicles. *Energy* 140, 1182–1197. <https://doi.org/10.1016/j.energy.2017.09.024>.
- Ganjisaffar, Y., Caruana, R., Lopes, C.V., 2011. Bagging Gradient-Boosted Trees for High Precision, Low Variance Ranking Models. In: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, Beijing, China. <https://doi.org/10.1145/2009916.2009932>.
- Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. *Neural Comput.* 4, 1–58. <https://doi.org/10.1162/neco.1992.4.1.1>.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv: 1806.00069 [cs.AI]*.
- Hao, X., Wang, H., Lin, Z., Ouyang, M., 2020. Seasonal effects on electric vehicle energy consumption and driving range: A case study on personal, taxi, and ridesharing vehicles. *J. Clean. Prod.* 249, 119403.
- Holland, S.P., Mansur, E.T., Muller, N.Z., Yates, A.J., 2018. Distributional Effects of Air Pollution from Electric Vehicle Adoption. *J. Assoc. Environ. Resour. Econ.* 6 (S1), S65–S94. <https://doi.org/10.1086/701188>.
- Hu, X., Chen, N., Wu, N., Yin, B., 2021. The Potential Impacts of Electric Vehicles on Urban Air Quality in Shanghai City. *Sustainability* 13 (2). <https://doi.org/10.3390/su13020496>.
- Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. *Int. J. Forecast.* 22 (4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- Indeed, I. (2022). [Indeed.com](https://www.indeed.com).
- Javid, R.J., Nejat, A., 2017. A comprehensive model of regional electric vehicle adoption and penetration. *Transp. Policy* 54, 30–42. <https://doi.org/10.1016/j.tranpol.2016.11.003>.
- Jenn, A., Springel, K., Gopal, A.R., 2018. Effectiveness of electric vehicle incentives in the United States. *Energy Policy* 119, 349–356. <https://doi.org/10.1016/j.enpol.2018.04.065>.
- Jung, H., 2020. Fuel economy of plug-in hybrid electric and hybrid electric vehicles: effects of vehicle weight, hybridization ratio and ambient temperature. *World Electric Vehicle J.* 11 (2). <https://doi.org/10.3390/wevj11020031>.
- Kaschub, T., Jochem, P., Fichtner, W., 2016. Solar energy storage in German households: profitability, load changes and flexibility. *Energy Policy* 98, 520–532. <https://doi.org/10.1016/j.enpol.2016.09.017>.
- Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Vaughan, J.W., 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3313831.3376219>.
- Kavadi, D.P., Patan, R., Ramachandran, M., Gandomi, A.H., 2020. Partial derivative Nonlinear Global Pandemic Machine Learning prediction of COVID 19. *Chaos Solitons Fractals* 139, 110056. <https://doi.org/10.1016/j.chaos.2020.110056>.
- Knez, M., Jereb, B., Obrecht, M., 2014. Factors influencing the purchasing decisions of low emission cars: A study of Slovenia. *Transp. Res. Part D: Transp. Environ.* 30, 53–61. <https://doi.org/10.1016/j.trd.2014.05.007>.
- Lévy, P.Z., Drossinos, Y., Thiel, C., 2017. The effect of fiscal incentives on market penetration of electric vehicles: A pairwise comparison of total cost of ownership. *Energy Policy* 105, 524–533. <https://doi.org/10.1016/j.enpol.2017.02.054>.
- Meckling, J., Nahm, J., 2019. The politics of technology bans: Industrial policy competition and green goals for the auto industry. *Energy Policy* 126, 470–479. <https://doi.org/10.1016/j.enpol.2018.11.031>.
- Meisel, S., Merfeld, T., 2018. Economic incentives for the adoption of electric vehicles: A classification and review of e-vehicle services. *Transp. Res. Part D: Transp. Environ.* 65, 264–287. <https://doi.org/10.1016/j.trd.2018.08.014>.
- Mersky, A.C., Sprei, F., Samaras, C., Qian, Z., 2016. Effectiveness of incentives on electric vehicle adoption in Norway. *Transp. Res. Part D: Transp. Environ.* 46, 56–68. <https://doi.org/10.1016/j.trd.2016.03.011>.
- Ng, A. Y. (1997). *Preventing "Overfitting" of Cross-Validation Data*. International Conference on Machine Learning.
- Noori, M., Gardner, S., Tatari, O., 2015. Electric vehicle cost, emissions, and water footprint in the United States: Development of a regional optimization model. *Energy* 89, 610–625. <https://doi.org/10.1016/j.energy.2015.05.152>.
- National Oceanic and Atmospheric Administration. (2022). <http://climate.gov>.
- Omahne, V., Knez, M., Obrecht, M., 2021. Social Aspects of Electric Vehicles Research—Trends and Relations to Sustainable Development Goals. *World Electric Vehicle J.* 12 (1), 15. <https://doi.org/10.3390/wevj12010015>.
- Onat, N.C., Kucukvar, M., Tatari, O., 2015. Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analysis in the United States. *Appl. Energy* 150, 36–49. <https://doi.org/10.1016/j.apenergy.2015.04.001>.
- OpenIntro. (2022). <http://www.openintro.org>.
- Pan, S., Roy, A., Choi, Y., Eslami, E., Thomas, S., Jiang, X., Gao, H.O., 2019. Potential impacts of electric vehicles on air quality and health endpoints in the Greater Houston Area in 2040. *Atmos. Environ.* 207, 38–51. <https://doi.org/10.1016/j.atmosenv.2019.03.022>.
- Pan, S., Fulton, L.M., Roy, A., Jung, J., Choi, Y., Gao, H.O., 2021. Shared use of electric autonomous vehicles: Air quality and health impacts of future mobility in the United States. *Renew. Sustain. Energy Rev.* 149, 111380. <https://doi.org/10.1016/j.rser.2021.111380>.
- Patt, A., Aplyn, D., Weyrich, P., van Vliet, O., 2019. Availability of private charging infrastructure influences readiness to buy electric cars. *Transp. Res. A Policy Pract.* 125, 1–7. <https://doi.org/10.1016/j.tra.2019.05.004>.
- Richardson, D.B., 2013. Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration. *Renew. Sustain. Energy Rev.* 19, 247–254. <https://doi.org/10.1016/j.rser.2012.11.042>.
- Sanguinetti, A., Salmon, K., Nicholas, M., Tal, G., Favetti, M., 2017. *Electric Vehicle Explorer. Lecture Notes in Computer Sciences*.
- Sovacool, B.K., Kester, J., Noel, L., de Rubens, G.Z., 2018. The demographics of decarbonizing transport: The influence of gender, education, occupation, age, and household size on electric mobility preferences in the Nordic region. *Glob. Environ. Chang.* 52, 86–100. <https://doi.org/10.1016/j.gloenvcha.2018.06.008>.
- Stokes, L.C., Breetz, H.L., 2018. Politics in the U.S. energy transition: Case studies of solar, wind, biofuels and electric vehicles policy. *Energy Policy* 113, 76–86. <https://doi.org/10.1016/j.enpol.2017.10.057>.
- Liberty Access Technologies. (2022). www.libertyaccesstechnologies.com.
- US Department of Energy. (2022). <http://energy.gov>.
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* 30, 79–82.
- World Media Group, LLC. (2022). <http://usa.com>.
- Xu, B., Sharif, A., Shahbaz, M., Dong, K., 2021. Have electric vehicles effectively addressed CO2 emissions? Analysis of eight leading countries using quantile-on-quantile regression approach. *Sustainable Production and Consumption* 27, 1205–1214. <https://doi.org/10.1016/j.spc.2021.03.002>.
- Zhang, Z., Jin, W., Jiang, H., Xie, Q., Shen, W., Han, W., 2017. Modeling heterogeneous vehicle ownership in China: A case study based on the Chinese national survey. *Transp. Policy* 54, 11–20. <https://doi.org/10.1016/j.tranpol.2016.10.005>.
- Zhao, H., Burke, A., 2014. An intelligent solar powered battery buffered EV charging station with solar electricity forecasting and EV charging load projection functions. *IEEE International Electric Vehicle Conference*.